

Substantiating Big Data in Health Care

NATHAN CORTEZ*

CONTENTS

I.	INTRODUCTION.....	61
II.	BIG DATA'S BREADTH	62
	A. <i>Breadth of Sources</i>	63
	B. <i>Breadth of Methods</i>	65
	C. <i>Breadth of Uses</i>	66
III.	CONCERNS, OLD AND NEW	71
IV.	EXISTING PARADIGMS	72
	A. <i>The Medical Product Paradigm</i>	73
	B. <i>The Medical Practice Paradigm</i>	78
	C. <i>The Medical Information Paradigm</i>	81
V.	CONTOURS OF A NEW PARADIGM?	81

I. INTRODUCTION

Mining large data sets for hidden connections promises to improve almost every sector of society, from business to government, academia, and private life. But some of the more compelling uses of “big data” are in health care. Predictive analytics might help us diagnose diseases more quickly, treat medical conditions more effectively, deploy scarce health resources more efficiently, and even improve the way we maintain our minds and bodies. Already, big data is being marshaled to these ends by nearly every major constituent of our health care system, including hospitals, insurers, practitioners,

* Callejo Endowed Professor, Associate Dean for Research, and Gerald J. Ford Research Fellow, SMU Dedman School of Law. I thank participants at events at Ohio State and Stanford for their comments and questions on earlier drafts, particularly Efthimios Parasidis and Jessica Roberts.

patients, researchers, product manufacturers, and government programs.

As data collection becomes more pervasive (and invasive), and as machine learning and analytical methods become more sophisticated, big data's predictive power will intensify. And as the big data industry matures, the companies that traffic in health-related big data will face competitive pressures to make more aggressive claims regarding what their analytics can predict. One can imagine, for example, hospital systems making speculative claims to treat cancers more effectively than competitors based on their use of proprietary data analytics, or insurers marketing data-driven plans that promise massive cost-savings. Already, patients and practitioners are inundated with claims by thousands of smartphone apps that promise to use data in novel ways to diagnose or manage scores of medical conditions. In short, our health system may become flooded with claims that big data and predictive analytics can solve our most pressing problems.

This article considers the level of evidence and substantiation that we should require of big data claims—focusing on “health claims,” or claims to diagnose, treat, or manage diseases or other medical conditions. Currently, there are three very different paradigms that might apply to such claims, depending on whether the predictions are cast as medical products, as medical practice, or merely as medical information. For example, should we treat predictive analytics akin to products like radiation treatment planning software, subject to controls on how it is manufactured and marketed? Or, is predictive analytics more akin to professional medical practice, and amenable to soft oversight by state licensors, practice guidelines, and medical specialty societies? Or is it most like medical information disseminated through journals, books, and the like, subject only to peer review?

I argue that because big data in health care is so opaque—both the data and the processing—big data claims may be uniquely difficult to substantiate, and thus call for a new paradigm. To date, scholars have considered new paradigms for big data privacy, research ethics, and the like. But this article considers what the contours of a new paradigm for evaluating big data health claims might look like, articulating key principles based on how predictive analytics departs from more traditional medicine.

II. BIG DATA'S BREADTH

Among the many challenges in substantiating big data health claims, three stand out. First is the volume and variety of data sources.

Second is the breadth, complexity, and fluidity of the methods for processing the data. Third is the breadth of different uses, some of which may require very different levels of substantiation.

A. *Breadth of Sources*

Almost everyone and everything that touches the U.S. health system contributes data to it. Big data in health care derives from patients, practitioners, hospitals, researchers, pharmacies, insurers, manufacturers, regulators, and countless other sources. Some sources are familiar, such as electronic medical records (EMRs), electronic health records (EHRs),¹ and insurance claims, including massive all-payer claims databases that collect medical, pharmacy, and dental claims from both public and private payers.² But big data also draws from records of prescriptions, diagnostic imaging results, laboratory tests (including genetic and genomic testing), outcomes reporting, and the many other records generated during health care delivery and payment. More recently, we have also begun to collect more continuous and granular data from mobile devices such as smartphones, wearables, ingestibles, implantables, and other sensors.³

These data sources are the product of perhaps hundreds of public and private sector initiatives that have pushed for greater transparency, accountability, and efficiency in health care. Some well-known programs have required great effort and expenses. For example, in 2008 the FDA launched its Sentinel Initiative, a national electronic system for monitoring safety problems with drugs, vaccines,

¹ EHRs are much broader than EMRs, which are merely “a digital version of the paper charts in the clinician’s office,” containing the patient’s medical and treatment history at that site. In contrast, the EHR is intended to span multiple providers and facilities, traveling with the patient. See Peter Garrett & Joshua Seidman, *EMR vs. EHR – What is the Difference?*, HEALTH IT BUZZ (Jan. 4, 2011), <https://www.healthit.gov/buzz-blog/electronic-health-and-medical-records/emr-vs-ehr-difference/> [<https://perma.cc/UC79-Y943>].

² Note, however, that in *Gobeille v. Liberty Mutual Ins. Co.*, 136 S.Ct. 936, 954 (2016), the Supreme Court invalidated a Vermont law that required reporting for its all-payer claims database as preempted by ERISA, the Employee Retirement Income Security Act of 1974.

³ See, e.g., WILLIAM HANSON, SMART MEDICINE: HOW THE CHANGING ROLE OF DOCTORS WILL REVOLUTIONIZE HEALTH CARE 153–54 (2011); ERIC TOPOL, THE CREATIVE DESTRUCTION OF MEDICINE: HOW THE DIGITAL REVOLUTION WILL CREATE BETTER HEALTH CARE 194–95 (2012); Nathan Cortez, *The Mobile Health Revolution*, 47 U.C. DAVIS L. REV. 1173, 1197–98 (2014) [hereinafter Cortez, *Mobile Health Revolution*].

biologics, and medical devices, relying on data from EHRs, insurance claims, and other heterogeneous sources.⁴ Since 2011, the federal government has given billions of dollars to Medicare and Medicaid providers that can demonstrate their “meaningful use” of EHRs in practice; with the idea that such use will improve clinical outcomes, reduce inefficiencies, and better engage patients in their own care.⁵ In 2013, the *Wall Street Journal* and the Center for Public Integrity convinced a court to overturn a 33 year-old injunction that prevented the Centers for Medicare and Medicaid (CMS) from releasing any federal payment data that identified individual providers, thus releasing Medicare payment data for roughly 825,000 practitioners.⁶ And today, the CMS websites *Hospital Compare*, *Physician Compare*, and *Nursing Home Compare* provide quality data on thousands of Medicare providers.⁷

These represent just a few of the high-profile efforts to gather and disseminate data in the health sector. The volume and variety of data sources continues to grow almost exponentially. In 2012, worldwide health data occupied around 500 petabytes of memory (about 10

⁴ U.S. FOOD & DRUG ADMIN. (FDA), FDA’S SENTINEL INITIATIVE, <https://www.fda.gov/Safety/FDAsSentinelInitiative/default.htm> [<https://perma.cc/SY6R-MHCQ>]. The Sentinel Initiative was created by the Food and Drug Administration Amendments Act of 2007 (FDAAA), Pub. L. No. 110-85, § 905, 121 Stat. 823, 944 (2007).

⁵ U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES (HHS), DATA AND PROGRAM REPORTS, CENTERS FOR MEDICARE AND MEDICAID SERVICES (CMS), <https://www.cms.gov/regulations-and-guidance/legislation/ehrincentiveprograms/dataandreports.html> [<https://perma.cc/4PZG-X7GC>] (reporting more than \$24 billion spent by the Medicare EHR Incentive Program and \$11.6 billion by the Medicaid EHR Incentive Program from 2011 through March 31, 2017). These payments derive from the Health Information Technology for Economic and Clinical Health Act (HITECH Act), as part of the American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, tit. XIII, §§ 13001-13424, 123 Stat. 115, 226-79 (2009) (codified as amended in scattered sections of the U.S. Code).

⁶ For a history detailing the litigation, see Erik Wemple, *Why Credit the Wall Street Journal on the Medicare Data Release?*, WASH. POST (Apr. 11, 2014), <https://www.washingtonpost.com/blogs/erik-wemple/wp/2014/04/11/why-credit-the-wall-street-journal-on-the-medicare-data-release/> [<https://perma.cc/QV5K-UG9J>]. For the court order freeing the data, see Order, Fla. Med. Ass’n v. Dep’t of Health, Educ., & Welfare, 947 F. Supp. 2d 1325 (M.D. Fla. 2013) (No. 3:78-cv-178-J-34MCR), <http://hr.cch.com/hld/FloridaMedicalAssociationvDeptofHealthEducationandWelfare.pdf> [<https://perma.cc/7A7J-YRQ4>].

⁷ For an analysis of these data efforts, see Nathan Cortez, *Regulation by Database*, 89 COLO. L. REV. (forthcoming 2017), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2951968.

billion four-drawer filing cabinets).⁸ But by 2020, health data will occupy roughly 50 times more, growing to 25,000 petabytes (500 billion cabinets).⁹ Moreover, even data traditionally situated outside the health sector—such as shopping patterns, GPS location data, web surfing habits, social media use, and the like—can be used to make important inferences about one's health. Now, even ostensibly non-health data can be “medically inflected.”¹⁰ As a result, any standards for substantiating big data health claims must account for the sheer volume and variety of data sources.

B. *Breadth of Methods*

The second complication for substantiating big data health claims is understanding the analytical methods used—no small feat. Predictive analytics, writ large, draws inferences from large data sets, relying on hypothesis-free data mining and inductive reasoning to uncover patterns.¹¹ Ziad Obermeyer and Ezekiel Emanuel explain that machine learning “approaches problems as a doctor progressing through residency might: by learning rules from data,” taking patient-level data to “sift through vast numbers of variables, looking for combinations that reliably predict outcomes.”¹² Nicholson Price describes the process as a computer system “trying a certain solution, evaluating the outcome, and then modifying that solution accordingly

⁸ Silvia Piai & Massimiliano Claps, *Bigger Data for Better Healthcare*, in IDC HEALTH INSIGHTS 3 (Sept. 2013), <http://www.intel.com/content/www/us/en/healthcare-it/solutions/documents/bigger-data-better-healthcare-idc-insights-white-paper.html> [<https://perma.cc/9GZ3-9B4S>] (click “Download the full report”).

⁹ *Id.*

¹⁰ Nicolas Terry, *Protecting Patient Privacy in the Age of Big Data*, 81 U. MISSOURI-KANSAS CITY L. REV. 1, 9 (2012). To truly appreciate the breadth of big data sources in health care, see Isaac Kohane, *Finding the Missing Link for Big Biomedical Data*, <https://dbmi.hms.harvard.edu/sites/g/files/mcu491/f/jvp140057f1-1050x855.png> [<https://perma.cc/4CPW-EJ9T>].

¹¹ For reasonably accessible descriptions of the different methods and how they evolved, see PEDRO DOMINGOS, *THE MASTER ALGORITHM: HOW THE QUEST FOR THE ULTIMATE LEARNING MACHINE WILL REMAKE OUR WORLD* (2015); PETER FLACH, *MACHINE LEARNING: THE ART AND SCIENCE OF ALGORITHMS THAT MAKE SENSE OF DATA* 361-62 (2012). For a more concise description, see W. Nicholson Price II, *Black Box Medicine*, 28 HARV. J.L. & TECH. 419, 432-34 (2014-15).

¹² Ziad Obermeyer & Ezekiel J. Emanuel, *Predicting the Future – Big Data, Machine Learning, and Clinical Medicine*, 375 NEW ENG. J. MED. 1216, 1217 (Sept. 29, 2016).

to improve future outcomes,” using as examples Pandora (music) and Netflix (movies).¹³ Of course, there are numerous variations of predictive analytics, including natural language processing, signal processing, topic modeling, pattern recognition, machine learning, deep learning, neural networks, and other advanced statistical methods.

If the breadth of methods is not daunting enough, consider that the methods themselves can be exceedingly complex and opaque, even to the programmers responsible for applying them.¹⁴ One, the relationships identified often remain hidden, even if the machine learning process is transparent and understood.¹⁵ Two, the methods continuously evolve, are adapted to different circumstances, and can be combined to amplify their predictive power, making each use a moving target.¹⁶ Three, the methods and applications may be proprietary and thus not available for scrutiny by third parties. Thus, in many instances, no one really understands or can explain the complex biological relationships identified by a predictive analytics program.¹⁷

C. Breadth of Uses

The third challenge for substantiating big data claims is the breadth of different uses. Big data is being marshaled for clinical decision support and patient care, including diagnosis, treatment, and chronic disease management. It is also a tool for hospitals and insurers (the latter somewhat problematically) to optimize triage and resource allocation decisions. Finally, big data is particularly compelling in biomedical research, including drug discovery and the development of novel diagnostic tools. Consider the following examples.

Physicians are making use of clinical decision support (CDS) systems that can turn vast amounts of data into actionable advice.¹⁸

¹³ Price, *supra* note 11, at 432.

¹⁴ *Id.* at 433.

¹⁵ *Id.*

¹⁶ See DOMINGOS, *supra* note 11, for a history of this evolution.

¹⁷ Price, *supra* note 11, at 434.

¹⁸ See, e.g., John D. Halamka, *Early Experiences with Big Data at an Academic Medical Center*, 33 HEALTH AFF. 1132, 1133 (July 2014).

Although the use of automated decision support is not at all new,¹⁹ the sheer volume of data available today—as well as the shift from rule-based systems modeled on experts to rules-generating machine learning²⁰—makes CDS systems much more powerful. A modern CDS system may be fed by millions of pages of electronic health records, laboratory results, vital sign readouts, prescription histories, radiology images, as well as vast libraries of medical literature, including best practice guidelines from expert societies. Without automation, the volume of information can overwhelm. Indeed, it can take an hour for a clinician to fully review just a single patient's EHR.²¹ But CDS systems can process these data to identify meaningful trends and events, and thus recommend in real-time a specific diagnostic test, a change in medication, or an intervention that improves the patient's outcome, conserves resources, or perhaps both.²² CDS thus allows clinicians to review a menu of evidence-based suggestions rather than the raw, underlying data.²³

At the institutional level, hospitals and health systems can use predictive analytics for triage, resource allocation, and quality enhancement efforts. For example, a single hospital might use big data to better manage high-cost patients, reduce hospital readmissions, triage clinical workflows to the most pressing cases, predict which patients' conditions might worsen and when (decompensation), predict adverse events, and better manage costly chronic conditions based on longitudinal data.²⁴ One model is Kaiser Permanente, which maintains 44 petabytes of EHR data for its roughly nine million patients—44-times more data than that contained in the entire Library of Congress.²⁵

¹⁹ Nathan Cortez, *Analog Agency in a Digital World*, in *FDA IN THE 21ST CENTURY: THE CHALLENGES OF REGULATING DRUGS AND NEW TECHNOLOGIES* 438, 439-40 (Holly Fernandez Lynch & I. Glenn Cohen, eds., 2015) [hereinafter Cortez, *Analog Agency*].

²⁰ Obermeyer & Emanuel, *supra* note 12.

²¹ Halamka, *supra* note 18, at 1133.

²² *Id.*

²³ *Id.*

²⁴ David W. Bates et al., *Big Data in Health Care: Using Analytics to Identify and Manage High-Risk and High-Cost Patients*, 33 *HEALTH AFF.* 1123, 1128 (2014).

²⁵ Caitlin M. Cusack et al., *The Future State of Clinical Data Capture and Documentation: A Report from AMIA's 2011 Policy Meeting*, 20(1) *J. AM. MED. INFORMATICS ASS'N* 134 (2013).

Kaiser's integrated EHR system has saved it roughly \$1 billion by reducing laboratory tests and physician office visits.²⁶ One ambitious estimate predicts that the use of big data in health care could save between \$300-450 billion in annual spending.²⁷ Moreover, new government reimbursement models that encourage patient outcome tracking and cost-conscious delivery, such as Accountable Care Organizations (ACOs),²⁸ may generate further data innovations.

Predictive analytics are also being incorporated into diagnostics. For example, the company Foundation Medicine markets "next-generation sequencing" (NGS) assays that can sequence cancer cells (located in solid tumors or circulating in the blood) and predict which therapies might best target the cancer based on its specific mutations and alterations.²⁹ The assay tests the entire coding sequences of 315 genes (plus introns from 28 additional genes) known to be related to cancer.

Another example is the Veterans Health Administration (VHA), which services over eight million patients through roughly 150 medical centers and 1,400 clinics. The VHA has been eager to adopt mobile health and big data tools to improve diagnosis and treatment.³⁰ One such use is the Durkheim Project, a collaboration between the VHA and Facebook, which tries to predict suicide risk by gathering data from veterans that choose to share information from

²⁶ Basel Kayyali, David Knott, & Steve Van Kuiken, *The Big Data Revolution in U.S. Health Care: Accelerating Value and Innovation*, MCKINSEY & CO. (Apr. 2013), <http://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care> [https://perma.cc/V2GK-LM3Y].

²⁷ PETER GROVES, BASEL KAYYALI, DAVID KNOTT & STEVE VAN KUIKEN, CTR. FOR U.S. HEALTH SYS. REFORM, *THE BIG DATA REVOLUTION IN HEALTH CARE: ACCELERATING VALUE AND INNOVATION* 8 (2013), https://www.ghdonline.org/uploads/Big_Data_Revolution_in_health_care_2013_McKinsey_Report.pdf [https://perma.cc/6P6N-X8KA].

²⁸ Patient Protection and Affordable Care Act of 2010, Pub. L. No. 111-148, § 3022, 124 Stat. 119, 395 (2010) (current version at 42 U.S.C. § 1395jjj (2010)).

²⁹ *What Is FoundationOne?*, FOUND. MED., <https://www.foundationmedicine.com/genomic-testing/foundation-one> [https://perma.cc/S28D-BN39].

³⁰ Christopher Wasden, *The Department of Veterans Affairs #mHealth Case Study*, HIMSS (Aug. 1, 2014), <http://www.himss.org/departement-veterans-affairs-mhealth-case-study> [https://perma.cc/2C32-DHDF].

their social media accounts and mobile devices, applying real-time prediction software.³¹

Yet another high-profile example is the use of IBM Watson's famous "cognitive computing system" in cancer care. In 2011, IBM partnered with the health insurer WellPoint to develop "a Watson-based diagnosis and treatment decision support system for oncology."³² In 2012, IBM then partnered with Memorial Sloan Kettering Cancer Center to process vast quantities of medical literature to generate recommendations for physicians, in theory replacing the traditional "tumor board."³³ Then in 2013, IBM and the MD Anderson Cancer Center announced that they would apply Watson to MD Anderson's large patient and research databases.³⁴ The idea was to combine data from diverse sources—patients, treatment records, research, and medical literature—and use Watson's artificial intelligence engine to recommend specific treatment options or clinical trials for individual patients.³⁵ As originally envisioned, the goal was to "transform how medicine will be practiced, by leveraging artificial intelligence," thus elevating the standard of cancer care

³¹ Jeff Bertolucci, *Big Data Project Analyzes Veterans' Suicide Risk*, INFO. WK. (July 9, 2013), <http://www.informationweek.com/big-data/big-data-analytics/big-data-project-analyzes-veterans-suicide-risk/d/d-id/1110668> [<https://perma.cc/ZPH2-Z9UT>]; Joachim Roski et al., *Creating Value in Health Care Through Big Data: Opportunities and Policy Implications*, 33 HEALTH AFF. 1115, 1116 (2014).

³² *Special Review of the Procurement Procedures Related to the M.D. Anderson Cancer Center Oncology Expert Advisor Project (2016-17)*, UNIV. OF TEX. SYS. AUDIT OFF., 2 (2016), <https://www.utsystem.edu/sites/default/files/documents/UT%20System%20Administration%20Special%20Review%20of%20Procurement%20Procedures%20Related%20to%20UTMDACC%20Oncology%20Expert%20Advisor%20Project/ut-system-administration-special-review-procurement-procedures-related-utmdacc-oncology-expert-advis.pdf> [<https://perma.cc/3P66-AMNK>] [hereinafter *University of Texas Audit*].

³³ Matthew Herper, *MD Anderson Benches IBM Watson in Setback for Artificial Intelligence in Medicine*, FORBES (Feb. 19, 2017), <http://www.forbes.com/sites/matthewherper/2017/02/19/md-anderson-benches-ibm-watson-in-setback-for-artificial-intelligence-in-medicine/#67aa13f37761> [<https://perma.cc/J7FV-5SZ7>]. Note, however, the emergence of "genomic tumor boards" in which traditional tumor boards incorporate genomic data, including recommendations from FoundationOne, into their discussions. See, e.g., Michelle L. McGowan et al., *A Rising Tide Lifts All Boats: Establishing a Multidisciplinary Genomic Tumor Board for Breast Cancer Patients with Advanced Disease*, 9 BMC MED. GENOMICS no. 71, 2016, at 4.

³⁴ Press Release, IBM, MD Anderson Taps IBM Watson to Power "Moon Shots" Mission Aimed at Ending Cancer, Starting with Leukemia (Oct. 18, 2013).

³⁵ *University of Texas Audit*, *supra* note 32, at 1.

worldwide.³⁶ The pilot first targeted leukemias, then expanded to lung cancer.³⁷

Other novel applications have also garnered widespread attention. Google Flu Trends was an effort to combine flu-related search engine queries with location data to predict in real time how cases of influenza might be spreading.³⁸ A later study showed that Google's model predicted twice as many cases of the flu than the model used by the Centers for Disease Control and Prevention (CDC).³⁹ But even after decommissioning Flu Trends in the United States, Google continues to monitor flu activity in 25 other countries, noting that "historically, national and regional estimates have been very consistent with traditional surveillance data collected by health agencies ... however, it is possible that future estimates may deviate from actual flu activity."⁴⁰ A similar location-based tool is Propeller Health (formerly Asthmapolis), a GPS asthma tracker that records symptoms and inhaler usage.⁴¹ The program merges its data with data from the CDC on pollen counts and other asthma triggers to develop personalized recommendations for users.

Finally, one of the more compelling uses of big data and predictive analytics in health care is for biomedical research, where researchers are querying massive heterogeneous data sets to mine for hidden relationships. Such queries have been facilitated by a few initiatives, including Informatics for Integrating Biology & the Bedside (I2B2), an

³⁶ *Id.* at 2.

³⁷ *Id.* at 3. The IBM-MD Anderson partnership met a disappointing end in 2017. When the project was audited in 2016, it had not yet been approved for human investigational or clinical use. In February 2017, MD Anderson announced that it was placing the project on hold, opening the program to other bidders, after an audit by the University of Texas System revealed that MD Anderson had paid over \$62 million on the project (not including internal resources) without much to show for it. See *supra* note 32. MD Anderson has engaged separate consultants to evaluate the scientific bases and functional capabilities of the system, but IBM defended Watson by noting that OEA's recommendations were 90% accurate when benchmarked with experts. See Herper, *supra* note 33.

³⁸ Jeremy Ginsberg et al., *Detecting Influenza Epidemics Using Search Engine Query Data*, 457 NATURE 1012, 1012 (2009).

³⁹ Declan Butler, *When Google Got Flu Wrong*, 494 NATURE 155, 155 (2013).

⁴⁰ Dawn Fallik, *For Big Data, Big Questions Remain*, 33 HEALTH AFF. 1111, 1112 n.6 (quoting Google Flu Trends, Frequently Asked Questions, at <http://www.google.org/flu Trends/about/faq.html> (last accessed May 30, 2014)).

⁴¹ Tim Maly, *Asthmapping: Smart Inhalers Track Usage to Save Lives*, WIRED (Nov. 1, 2012), <https://www.wired.com/2012/11/asthmapolis/> [<https://perma.cc/5PF8-2JP4>].

open-source software tool developed at Harvard to help researchers mine clinical data,⁴² and Query Health, a broad methodology for querying hospital big data developed by the Office of National Coordinator for Health IT (ONC) within the U.S. Department of Health and Human Services (HHS).⁴³

III. CONCERNS, OLD AND NEW

The use of big data and predictive analytics in health care raises a number of novel legal and ethical concerns. For example, can insurers use predictive analytics to offer customized plans without violating laws that ban health status underwriting?⁴⁴ Does big data offer a stealth way for insurers to identify preexisting conditions and other health status factors?⁴⁵ Do patients have to give informed consent before their personal data are used in predictive models, or when they are subject to treatment decisions based on automated support?⁴⁶ Will hospitals or practitioners face liability for relying on a predictive model that errs, or for *not* following a valid recommendation?⁴⁷ What dangers lurk when big data practices fall outside the scope of health privacy laws like HIPAA? How can we detect and prevent errors caused by complex, opaque computer code? How does human-computer interaction skew clinical decision-making and judgment?

⁴² Introduction, I2B2: INFORMATICS FOR INTEGRATING BIOLOGY & THE BEDSIDE, <https://www.i2b2.org/about/intro.html> [<https://perma.cc/QKR2-RYHF>]; Griffin M. Weber et al., *The Shared Health Research Information Network (SHRINE): A Prototype Federated Query Tool for Clinical Data Repositories*, 16 J. AM. MED. INFORMATICS ASS'N 624, 625 (2009).

⁴³ Doug Fridsma, *Join Query Health in Developing National Standards for Population Queries*, HEALTHIT (Sept. 23, 2011), <https://www.healthit.gov/buzz-blog/from-the-onc-desk/queryhealth/> [<https://perma.cc/R2DH-J4P3>].

⁴⁴ Patient Protection and Affordable Care Act of 2010, *supra* note 28, at § 1201 (current version at 42 U.S.C. §§ 300gg-1, 300gg-2 (2010)).

⁴⁵ *Id.*

⁴⁶ I. Glenn Cohen et al., *The Legal and Ethical Concerns that Arise from Using Complex Predictive Analytics in Health Care*, 33 HEALTH AFF. 1139, 1143 (2014) (arguing that obtaining informed consent in such circumstances would be unworkable).

⁴⁷ *Id.* at 1144; Sharona Hoffman et al., *E-Health Hazards: Provider Liability and Electronic Health Record Systems*, 24 BERKELEY TECH. L.J. 1523, 1548, 1553 (2009) (arguing that clinicians and physicians can be liable for use faulty predictive models or ignoring recommendations); Halamka, *supra* note 18, at 1136-37 (arguing that data quality and inconsistencies pose liability problems for users of predictive models).

Some of these concerns are indeed novel, but many are not. In previous work, I have traced how contemporary handwringing over computerized medicine echoes the earliest concerns raised in the 1970s, when many potential uses were first conceived.⁴⁸ In addition to the concerns above, policymakers have long worried about regulating computerized medicine too early, too stringently, or too clumsily, in a way that failed to account for the unique nature of software.⁴⁹ Over the decades, observers have also worried about regulators that lack expertise overseeing an industry (software) that has largely been spared from federal regulation.⁵⁰ Finally, a persistent concern is that *any* regulation of computerized medicine would necessarily constitute regulation of medical knowledge itself.⁵¹

There is a growing literature that wrestles with these questions in depth. But here, I want to isolate a discrete challenge—how to substantiate health claims made by predictive analytics. The types of claims I have in mind are not claims to improve efficiency or resource allocation or research findings, but claims that a predictive analytics program will improve patient outcomes. For example, what kind of substantiation should we require of a mobile application that claims to distinguish ordinary moles from skin cancers based on pattern recognition and deep learning? Or machine learning software that claims to identify suicide risks among veterans? Or an artificial intelligence engine that claims to offer targeted therapies to cancer patients based on customized recommendations? What kinds of evidence, and how much of it, should we require?

IV. EXISTING PARADIGMS

The answers depend, to a large extent, on how we characterize predictive analytics in health care. Three existing paradigms might lend themselves here, though none fit perfectly. If we characterize predictive analytics as a medical product, akin to a drug or device, then purveyors will be required by the U.S. Food and Drug Administration (FDA) and Federal Trade Commission (FTC) to

⁴⁸ Cortez, *Analog Agency*, *supra* note 19, at 441, 447-51. For an early prediction that computers would gradually replace physicians, see JERROLD S. MAXMEN, *THE POST-PHYSICIAN ERA: MEDICINE IN THE 21ST CENTURY* 93 (David Mechanic ed., 1976).

⁴⁹ See Cortez, *Analog Agency*, *supra* note 19, at 447-51.

⁵⁰ *Id.* at 449-50.

⁵¹ *Id.* at 450-51.

substantiate their claims with valid scientific evidence, possibly clinical trials. If we characterize predictive analytics as a form of medical practice, then perhaps we treat the algorithm like a physician and require some sort of threshold license before clinical use and subject the algorithm to a reasonable standard of care? If we characterize predictive analytics as medical knowledge or information, then perhaps we leave traditional regulation out of it, and rely on peer review or some equivalent?

A. The Medical Product Paradigm

Characterizing predictive analytics as medical products would require claims to be substantiated in the same way claims for drugs, devices, and other medical products are substantiated. There is not, of course, a single evidentiary standard for all product claims; the standard varies depending on the type of product, the novelty of its claims, and the agency asserting jurisdiction.

For example, the FTC oversees health claims in advertising and marketing,⁵² and requires substantiation through “competent and reliable scientific evidence.”⁵³ This means that the evidence, “when considered in light of the entire body of relevant and reliable scientific evidence, is sufficient to substantiate that the representation is true.”⁵⁴ Claims must have substantiation before they are made, though the FTC does not require preapproval. But for certain health claims, such as claims made about specific diseases, the FTC can require randomized, double-blinded, placebo-controlled clinical trials, as it did for POM Wonderful’s claims that its juice products could help treat or prevent heart disease, cancer, and other conditions.⁵⁵ Although the FTC does not always require randomized trials, the agency has filed several complaints against mobile app developers

⁵² Although the F.T.C. has responsibility for over 70 separate statutes, see FED. TRADE COMM’N., FEDERAL TRADE COMMISSION ACT: INCORPORATING U.S. SAFE WEB ACT AMENDMENTS AS OF 2006, https://www.ftc.gov/sites/default/files/documents/statutes/federal-trade-commission-act/ftc_act_incorporatingus_safe_web_act.pdf [<https://perma.cc/78HX-9VTE>], we are primarily concerned here with sections 5 and 12 of the FTC Act, which prohibit unfair or deceptive trade practices and false advertising. See 15 U.S.C. §§ 45, 52.

⁵³ POM Wonderful v. F.T.C., 777 F.3d 478, 489 (D.C. Cir. 2015), *cert. denied*, 136 S. Ct. 1839 (2016).

⁵⁴ *Id.*

⁵⁵ *Id.*

making dubious health claims, such as an app claiming to improve vision, an app claiming to accurately measure blood pressure, and apps claiming to detect melanomas using smartphone imaging.⁵⁶ Thus, predictive analytics programs claiming to diagnose, treat, or manage specific health conditions must be able to produce competent and reliable scientific evidence to substantiate their claims, though the FTC has yet to articulate in many cases precisely what such evidence might look like.

Health claims are also subject to FDA oversight if the product qualifies as a medical “device.”⁵⁷ Over the last few decades, there has been considerable confusion over when software qualifies as a “device” subject to FDA jurisdiction.⁵⁸ But in the last few years, both the FDA and Congress have tried to provide more clarity,⁵⁹

⁵⁶ Carrot Neurotechnology, Docket No. C-4567 (F.T.C., Feb. 22, 2016), <https://www.ftc.gov/system/files/documents/cases/160223carrotneurodo.pdf> [<https://perma.cc/728F-BTHW>]; FTC v. Aura Labs, No.8:16-cv-2147-DOC (C.D. Cal. Dec. 9, 2016), https://www.ftc.gov/system/files/documents/cases/161212_aura_labs_final_order.pdf [<https://perma.cc/7YQF-XQR4>]; Health Discovery, (F.T.C., 2015), <https://www.ftc.gov/system/files/documents/cases/complaint.pdf> [<https://perma.cc/6PNF-7HBS>]; FTC, FTC CRACKS DOWN ON MARKETERS OF “MELANOMA DETECTION” APP (Feb. 23, 2015), <https://www.ftc.gov/news-events/press-releases/2015/02/ftc-cracks-down-marketers-melanoma-detection-apps> [<https://perma.cc/VK58-C7DN>]; FTC, MOBILE HEALTH APP DEVELOPERS: FTC BEST PRACTICES, <https://www.ftc.gov/tips-advice/business-center/guidance/mobile-health-app-developers-ftc-best-practices> [<https://perma.cc/CQ63-YEXN>]; see generally, FTC, HEALTH CLAIMS, <https://www.ftc.gov/tips-advice/business-center/advertising-and-marketing/health-claims> [<https://perma.cc/AW5Y-76Z4>].

⁵⁷ Federal Food, Drug, and Cosmetic Act § 201(h) (2012); 21 U.S.C. § 321(h) (2012).

⁵⁸ See, e.g., Cortez, *Mobile Health Revolution*, *supra* note 3; Cortez, *Analog Agency*, *supra* note 19; Nathan Cortez, I. Glenn Cohen, & Aaron S. Kesselheim, *FDA Regulation of Mobile Health Technologies*, 371 N. ENG. J. MED. 372, 374 (2014).

⁵⁹ Food and Drug Administration Safety and Innovation Act (FDASIA) § 618, Pub. L. No. 112-144, 126 Stat. 1063 (2012); U.S. Food & Drug Admin., FDASIA HEALTH IT REP. (April 2014), <https://www.fda.gov/downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDRH/CDRHReports/UCM391521.pdf> [<https://perma.cc/AZL2-NVMR>]; U.S. Food & Drug Admin., MOBILE MEDICAL APPLICATIONS: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (Feb. 9, 2015), <https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM263366.pdf> [<https://perma.cc/U95R-WM5H>]. Note that the FDASIA Health IT working group recommended against FDA regulation of data analytics, see, DW Bates, DRAFT FDASIA COMMITTEE REP., http://www.healthit.gov/facas/sites/faca/files/FDASIARRecommendationsDraftto30913_v2.pdf [<https://perma.cc/R7LD-FB2E>], unless used for “strong” clinical decision support or used in regulated medical devices. FDASIA HEALTH IT REP., *supra* note 59.

culminating in late 2016, when Congress passed the 21st Century Cures Act.⁶⁰ The Act amended the statutory definition of “device” to exempt from FDA jurisdiction certain types of software—including software used for administrative purposes, wellness and lifestyle purposes, patient record purposes, and some clinical decision support purposes.⁶¹

The new definition includes some interesting wrinkles for predictive analytics. First, it defines administrative software as software intended for processing “information about patient populations, admissions, practice and inventory management, analysis of historical claims data to predict future utilization or cost-effectiveness, determination of health benefit eligibility, population health management, and laboratory workflow.”⁶² Thus, predictive analytics used for hospital triage, resource allocation, or workflows would not, by definition, fall within FDA jurisdiction.

Second, the new definition exempts software intended to transfer, store, or display data from patient records, laboratory tests, or other device inputs, *unless* it is “intended to interpret or analyze” these data.⁶³ As such, predictive analytics that generate customized treatment recommendations might qualify as medical “devices” subject to FDA review.

Third, a closely related exemption carves out from the definition of “device” clinical decision support software unless the professional using it cannot “independently review” the basis for its recommendations and the professional is intended to rely on the recommendation as a primary point for diagnosing or treating a specific patient.⁶⁴ This standard recalls the FDA’s 1987 policy that exempted from agency scrutiny decision support software, but only if it allowed time for “competent human intervention,” meaning that “clinical judgment and experience can be used to check and interpret a system’s output” before “any impact on human health.”⁶⁵ Like the

⁶⁰ 21st Century Cures Act, Pub. L. No. 114-255, § 3060; 130 Stat. 1033 (2016).

⁶¹ *Id.* at § 3060(a) (codified at 21 U.S.C. § 360j(o)).

⁶² *Id.* at § 3060(a) (codified at 21 U.S.C. § 360j(o)(1)).

⁶³ *Id.* at § 3060(a) (codified at 21 U.S.C. §§ 360j(o)(1)(C), (D)).

⁶⁴ *Id.* at § 3060(a) (codified at 21 U.S.C. § 360j(o)(1)(E)).

⁶⁵ See Cortez, *Mobile Health Revolution*, *supra* note 3, at 1220 (discussing the FDA’s 1987 Draft Policy Guidance for Regulation of Computer Products, 52 Fed. Reg. 36,104 (Sept. 25, 1987), and the FDA’s 1989 Draft Policy for the Regulation of Computer Products (Nov. 13, 1989), available at 1989 WL 1178702).

earlier FDA standard, the 21st Century Cures Act overlooks modern thinking on human-computer interaction (HCI) and automation bias, which teaches that we are dangerously predisposed to trust computer-generated advice, even when we have reason to suspect that it errs.⁶⁶

Fourth, the new definition of “device” in the Cures Act also declines to exempt software that processes data from medical images, in vitro diagnostics, or signal and pattern acquisition systems⁶⁷—rich data sources that many hope will populate artificial intelligence engines for years to come.

Finally, notwithstanding the broad exemptions, the FDA can, through notice-and-comment in the *Federal Register*, assert jurisdiction over software that would otherwise be exempt if the agency finds that it “would be reasonably likely to have serious adverse health consequences.”⁶⁸ This may prove to be an important regulatory backstop when a predictive analytics program that initially falls outside FDA jurisdiction later evolves, increasing both its predictive powers and perhaps its ambitions (and intended uses).

Thus, the question of whether the FDA can regulate claims made by a predictive analytics program is an exercise of complex, somewhat tedious statutory construction. If the program is subject to FDA oversight, the agency will expect marketing claims to be truthful and not misleading, which again may require substantiation via clinical trials. Or, perhaps over time, if a predictive analytics program is “substantially equivalent” to an older predicate on the market, the FDA might clear the product for marketing based on a showing that it is at least as safe and effective as the previous, legally marketed device.⁶⁹ One important barrier to a finding of “substantial equivalence,” however, is when the newer device has a different intended use than the predicate.⁷⁰ For example, a machine learning program designed to target one disease would not be substantially equivalent if repurposed to target a different disease.

The FDA also expects software under its jurisdiction to satisfy validation requirements, which help ensure that the software operates

⁶⁶ *Id.* at 1226-27.

⁶⁷ 21st Century Cures Act, *supra* note 60, at § 3060(a) (codified at 21 U.S.C. § 360j(o)(1)(E)).

⁶⁸ *Id.* at § 3060(a) (codified at 21 U.S.C. § 360j(o)(3)).

⁶⁹ 21 C.F.R. § 807.92(a)(3) (2017).

⁷⁰ *Id.* at § 807.81(a)(3).

according to user needs and intended uses.⁷¹ Thus, the FDA might expect predictive analytics programs to validate the model's sensitivity, predictive values (positive and negative), c-statistics, and other performance metrics.⁷² Scholars have called for more rigorous validation of predictive analytics models if the risks to patients is relatively high.⁷³ For example, "if the model's predictions could direct clinicians to withhold interventions recommended by current evidence-based guidelines," the model might warrant outcomes and analysis plans, peer review, and other robust validation methods.⁷⁴ If the risks to patients are low, perhaps less rigorous validation, such as post-hoc outcomes comparisons, might suffice.⁷⁵

The problem with applying the FDA's product-based model to predictive analytics is that it doesn't fit particularly well. The FDA's medical device framework was developed in 1976 with more traditional, tangible devices in mind. Few devices then were controlled by, or even incorporated, software. At the time, only science fiction writers speculated about the types of technologies used today.⁷⁶ But by 2006, more than half of all medical devices on the U.S. market incorporated software.⁷⁷ This shift occurred with very little change in the FDA's statutory authority or approach to regulating medical devices.⁷⁸ Today, then, medical devices are vastly different than in 1976, but the FDA's basic approach is not. It is thus unsurprising that developers are frustrated with the time it takes FDA to clear both new

⁷¹ See U.S. FOOD & DRUG ADMIN., GENERAL PRINCIPLES OF SOFTWARE VALIDATION; FINAL GUIDANCE FOR INDUSTRY AND FDA STAFF (2002); U.S. FOOD & DRUG ADMIN., GUIDANCE FOR THE CONTENT OF PREMARKET SUBMISSIONS FOR SOFTWARE CONTAINED IN MEDICAL DEVICES (2005); 21 C.F.R. § 820.3(z) (2017).

⁷² Cohen et al., *supra* note 46, at 1143.

⁷³ *Id.* at 1142.

⁷⁴ *Id.*

⁷⁵ *Id.*

⁷⁶ Cortez, *Mobile Health Revolution*, *supra* note 3, at 1175-76 (recounting the famous medical Tricorder from Star Trek).

⁷⁷ Kevin Fu, Trustworthy Medical Device Software, Institute of Medicine Workshop on the FDA's 510(k) Process at 35 Years (2011) (prepublication draft), <https://spqr.eecs.umich.edu/papers/fu-trustworthy-medical-device-software-IOM11.pdf> [<https://perma.cc/45UL-PNHQ>].

⁷⁸ Cortez, *Analog Agency*, *supra* note 19, at 450-51.

products and modifications to existing products.⁷⁹ Moreover, the FDA's validation requirements might be particularly difficult to apply to predictive analytics models that are non-transparent, such as when the model relies on artificial intelligence or machine learning that evolves frequently—so-called “black box medicine.”⁸⁰ These complications make predictive analytics seem more like medical practice than medical products.

B. *The Medical Practice Paradigm*

Is it more appropriate to characterize predictive analytics as a form of medical practice? Consider an example. In January 2017, Stanford researchers demonstrated that a deep learning program could diagnose skin cancers on equal footing with dermatologists.⁸¹ Using a Google-based deep convolutional neural network that was pre-trained on 1.28 million images for 1,000 different object categories, the researchers then trained the algorithm on 129,450 images of skin lesions labeled by dermatologists, containing a taxonomy of 2,032 different diseases.⁸² The researchers then tested the deep neural network against 21 board-certified dermatologists, using 370 high-quality images.⁸³ The results, published in *Nature*, showed that the deep neural network matched the dermatologists, using a sensitivity-specificity curve to calculate the proportion of malignant and benign lesions classified correctly.⁸⁴ The researchers were optimistic that use of the deep learning program—including its use on mobile devices—could greatly improve early detection of skin cancers, though they emphasized that “rigorous prospective validation” would be required before use by clinicians or patients.⁸⁵

⁷⁹ Cortez et al., *supra* note 58, at 373-75.

⁸⁰ Cohen et al., *supra* note 46, at 1142-43; Price, *supra* note 11, at 421.

⁸¹ Andre Esteva et al., *Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks*, 542 NATURE 115, 115 (Feb. 2, 2017).

⁸² *Id.*

⁸³ *Id.* at 115, 118.

⁸⁴ *Id.* at 117-18.

⁸⁵ Taylor Kubota, *Deep Learning Algorithm Does as Well as Dermatologists in Identifying Skin Cancer*, STANFORD NEWS (Jan. 25, 2017), <http://news.stanford.edu/2017/01/25/artificial-intelligence-used-identify-skin-cancer/> [<https://perma.cc/2L7H-8MK3>].

What kind of substantiation should we require for an algorithm like this?

Regulating such an algorithm like medical practice would have bizarre but provocative implications. For example, there is little equivalent for training algorithms the way we train dermatologists—through four years of medical school, a one-year internship in general practice, a three-year residency practicing under experienced dermatologists, and perhaps a one-year fellowship, not to mention passing the obligatory state medical boards and certification by the American Academy of Dermatology. All are predicates to calling oneself a dermatologist. Of course, the training and testing does not stop there. The practitioner must also maintain all licenses and certifications, which requires occasional continuing medical education (CME) and perhaps more tests. Finally, the practitioner may have to obtain staff privileges at one or more hospitals and thus be subject to hospital credentialing, peer review, and other quality control efforts. During these years, the student-turned-dermatologist would examine untold numbers of skin lesions, ideally developing a highly-trained intuition that is refined over time by even more training and experience.

The algorithm, by contrast, is trained almost by brute force. The skin cancer algorithm above “requires no hand-crafted features: it is trained end-to-end directly from image labels and raw pixels,” fed by well over a million pre-training and training images.⁸⁶ Yet, despite the different training methods, both the dermatologist and the algorithm learn “rules” for diagnosing cancers through data—albeit different kinds of data.⁸⁷

Although the algorithm matched the performance of the 21 dermatologists, neither group could classify every image accurately. Indeed, it might be unrealistic or unfair to expect infallibility from an algorithm and thus hold it to a higher standard than physicians, whose fallibility and errors are well documented.⁸⁸ Thus, if we characterize predictive analytics as a form of medical practice, perhaps we validate its performance against physician performance as a barometer? This approach is both intuitive and symmetrical.

⁸⁶ Esteva et al., *supra* note 81, at 115.

⁸⁷ Ziad Obermeyer & Ezekiel J. Emanuel, *Predicting the Future – Big Data, Machine Learning, and Clinical Medicine*, 375 NEW ENG. J. MED. 1216, 1217 (2016).

⁸⁸ For the iconic study, see INSTITUTE OF MEDICINE, *TO ERR IS HUMAN: BUILDING A SAFER HEALTH SYSTEM* 39 (2000), <https://www.nap.edu/read/9728/chapter/1> [<https://perma.cc/HP28-ED8T>].

However, the symmetry is distorted when accounting for marketing claims. Physicians usually avoid making the more strident claims that emerge from corporate sales and marketing departments, deterred perhaps by professional norms, the threat of malpractice liability, or both (although companies, in theory, are also constrained by the threat of product liability and unfair trade practice laws).⁸⁹ Physicians cannot, and thus generally do not, guarantee results,⁹⁰ in contrast to some products. Thus, although it seems perfectly natural to measure the performance of predictive analytics against the performance of physicians, it seems unnatural to expect companies to restrain their marketing claims the same way physicians do. After all, the familiar disclaimer in advertisements that “Results may vary” is only necessary when the advertisement implies the very opposite.

The answer, perhaps, is to subject predictive analytics to the same legal duty of care we expect from practitioners, or hospitals, or from medical products⁹¹—though again, the duty will vary depending on whether we characterize the technology as medical practice, as a medical product, or as a medical enterprise such as a hospital. Regardless, it may prove exceedingly difficult to determine when an algorithm falls below the standard of care (in the case of medical malpractice or corporate negligence claims) or is defective (in the case of product liability claims).⁹² On what basis could we call a machine learning algorithm or a deep neural network “sub-standard” or “defective”? If it underperforms practitioners? If it underperforms other predictive analytics programs?

In sum, characterizing predictive analytics as a form of medical practice has intuitive appeal because it relies on a standard of reasonable care, based on what we expect of contemporary practitioners. But there may be no good analog in medical practice regulation for substantiating health claims.

⁸⁹ See, e.g., Nicolas P. Terry & Lindsey F. Wiley, *Liability for Mobile Health and Wearable Technologies*, 25 ANNALS HEALTH L. 62, 70 (2016); Nathan Cortez, *The Evolving Law and Ethics of Digital Health*, in DIGITAL HEALTH: SCALING HEALTHCARE TO THE WORLD (forthcoming 2017) [hereinafter Cortez, *The Evolving Law*].

⁹⁰ Terry & Wiley, *supra* note 89, at 70.

⁹¹ See, *id.*; see Cortez, *The Evolving Law*, *supra* note 89 (describing medical licensing, medical malpractice, enterprise liability, and product liability as it applies to digital health products).

⁹² See Terry & Wiley, *supra* note 89, at 70-71; See Cortez, *The Evolving Law*, *supra* note 89.

C. *The Medical Information Paradigm*

Is it more appropriate to characterize predictive analytics merely as medical information, akin to a medical journal article, a reference text, or even an interactive web site? Policymakers have long worried that regulating medical software would equate to regulating medical knowledge—which of course is unregulated in any traditional sense.⁹³ The closest thing to quality control we have for medical textbooks, journal articles, and reference sources is peer review and professional editing. Some observers even worry that government regulation of medical software might violate First Amendment rights insofar as it restricts the dissemination of medical information as speech.⁹⁴ Indeed, the government's core relationship with medical information is to subsidize it, not regulate it.⁹⁵

The clear implication of characterizing predictive analytics as medical information, then, is to effectively exempt it from any oversight. This paradigm may be appropriate for predictive analytics used for research purposes, but it would seem unfair to medical professionals, medical products, and medical institutions that are subject to much more stringent requirements, based on public health and safety rationales. Eventually, the more we begin to rely on predictive analytics to make clinical decisions, the more absurd it seems to characterize it as mere medical information, akin to a medical reference text.

V. CONTOURS OF A NEW PARADIGM?

If predictive analytics does not fit well into existing frameworks governing medical products, medical professionals, or medical information, what would a more appropriate framework look like? There is a fair bit of skepticism of traditional regulation by traditional regulators; conversely, there is a fair bit of optimism in private certifiers and technology-enabled intermediation.⁹⁶ The FDA itself, in

⁹³ Cortez, *Analog Agency*, *supra* note 19, at 450–51.

⁹⁴ *Id.*

⁹⁵ The best examples here may be the Human Genome Project and the BRAIN Initiative, through which the federal government has spent billions subsidizing research, in addition to the billions spent by the NIH (National Institutes of Health) each year.

⁹⁶ FDASIA Health IT Report, *supra* note 59; Price, *supra* note 11; U.S. Food and Drug Admin., FDASIA Health IT Report: Proposed Strategy and Recommendations for a Risk-Based Framework (Apr. 2014),

fact, announced in 2017 that it was going to experiment with the use of private certifiers for digital health products.⁹⁷ Calls for Congress to authorize more tailored regulation by the FDA⁹⁸ were met instead with relatively tepid clarifications in the 21st Century Cures Act that nevertheless failed to address longstanding confusion over clinical decision support and other big picture questions. Still, if Congress ever were to contemplate a new framework, it should consider the following principles:

First, there is widespread agreement that not all medical software requires regulation, such as administrative software, software targeted at general fitness and well-being, and other low-risk software; the focus instead should be on software that claims to help diagnose or treat patients and presents moderate to high risks.⁹⁹ Nevertheless, given how easy it can be to modify software and add functions, policymakers should anticipate that otherwise benign software can easily evolve or be merged with other programs, thus breaching regulatory boundaries. There is also evidence that developers frequently design around the jurisdiction of regulators like the FDA.¹⁰⁰ Thus, although there is consensus that low-risk software warrants

<https://www.fda.gov/downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDRH/CDRHReports/UCM391521.pdf> [<https://perma.cc/33MZ-V6CT>]; W. Nicholson Price II, *Black Box Medicine*, 28 HARV. J. L. & TECH. 419, 432-34 (2015); Dov Greenbaum, *Avoiding Over-Regulation in the Medical Internet of Things*, in BIG DATA, HEALTH LAW, AND BIOETHICS (Holly Fernandez Lynch, I. Glenn Cohen, Effy Vayena, & Urs Gasser eds.) (forthcoming, Cambridge 2018); Marcus Comiter, *Data Policy for Internet of Things Healthcare Devices: Aligning Patient, Industry, and Privacy Goals in the Age of Big Data*, in BIG DATA, HEALTH LAW, AND BIOETHICS (Holly Fernandez Lynch, I. Glenn Cohen, Effy Vayena, & Urs Gasser eds.) (forthcoming, Cambridge 2018). For a more critical assessment of private certifiers and intermediaries, see Frank Pasquale, *Private Certifiers and Deputies in American Health Care*, 92 N.C. L. REV. 1661, 1673, 1676 (2014).

⁹⁷U.S. FOOD AND DRUG ADMIN., DIGITAL HEALTH INNOVATION ACTION PLAN (July 2017), <https://www.fda.gov/downloads/MedicalDevices/DigitalHealth/UCM568735.pdf> [<https://perma.cc/F7QN-DAQ6>]; Nathan G. Cortez, Nicolas P. Terry & I. Glenn Cohen, *Questions About the FDA's New Framework for Digital Health*, HEALTH AFFAIRS BLOG (Aug. 16, 2017), <http://healthaffairs.org/blog/2017/08/16/questions-about-the-fdas-new-framework-for-digital-health/> [<https://perma.cc/6CWV-Z5H8>].

⁹⁸ Cortez et al., *supra* note 58, at 375; Cortez, *Analog Agency*, *supra* note 19, at 451-53.

⁹⁹ See, e.g., U.S. Food and Drug Admin., *supra* note 96; 21st Century Cures Act, *supra* note 60, at § 3060.

¹⁰⁰ Nicolas P. Terry, *Big Data and Regulatory Arbitrage in Health Care*, in BIG DATA, HEALTH LAW, AND BIOETHICS (Holly Fernandez Lynch, I. Glenn Cohen, Effy Vayena, & Urs Gasser eds.) (forthcoming, Cambridge 2018); Cortez, *Mobile Health Revolution*, *supra* note 3, at 1187-88 (discussing disclaimers).

little attention, the fluid and malleable nature of software requires some flexibility for regulators.¹⁰¹

Second, any new framework should consider the software's intended users. There is some agreement that programs targeting patient users should receive more scrutiny than programs targeting more sophisticated users, such as health practitioners and institutions.¹⁰² Thus, we might logically expect more scientific substantiation for skin cancer-detecting software when it is targeted at smartphone users than we would for software targeted at dermatologists.¹⁰³ Still, history tells us that even sophisticated users can benefit from quality screening.¹⁰⁴ For example, practitioners are often bewildered by the sheer number of health apps available.¹⁰⁵ Moreover, even world-class institutions can struggle to incorporate predictive analytics. A recent example is MD Anderson's failed partnership with IBM Watson, which was the subject of a scathing audit by the University of Texas showing that MD Anderson had spent over \$62 million on the project, with little to show for it.¹⁰⁶ Any new

¹⁰¹ Indeed, the backstop provided by the 21st Century Cures Act seems to provide FDA the ability to assert jurisdiction over high-risk software that it otherwise might not regulate, so long as the agency uses notice-and-comment procedures. See 21st Century Cures Act, *supra* note 60.

¹⁰² See, U.S. FOOD AND DRUG ADMIN., MOBILE MEDICAL APPLICATIONS: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (Feb. 9, 2015), <https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM263366.pdf> [<https://perma.cc/7R6S-4D8C>].

¹⁰³ For examples of the former, see, FTC, FTC CRACKS DOWN ON MARKETERS OF "MELANOMA DETECTION" APPS (2015), <https://www.ftc.gov/system/files/documents/cases/complaint.pdf> [<https://perma.cc/KW5A-PZZC>].

¹⁰⁴ For an interesting parallel, consider that one of the major rationales for not regulating mortgage-backed securities, credit default swaps, and other sophisticated collateralized debt instruments was that the transactions involved highly sophisticated parties that would fully understand the risks and benefits. That proved to be a dangerous assumption leading up to the 2008 financial crisis. See FIN. CRISIS INQUIRY COMM'N, FINANCIAL CRISIS INQUIRY REPORT xxiv-xxv, 47 (Jan. 2011); Nathan Cortez, *Regulating Disruptive Innovation*, 29 BERKELEY TECH. L.J. 173, 183-84 (2014).

¹⁰⁵ Cortez et al., *supra* note 58.

¹⁰⁶ See Kubota, *supra* note 85. See UNIV. OF TEX. SYS. AUDIT OFFICE, SPECIAL REVIEW OF THE PROCUREMENT PROCEDURES RELATED TO THE M.D. ANDERSON CANCER CENTER ONCOLOGY EXPERT ADVISOR PROJECT (2016-17), at 2,

paradigm should carefully consider whether and how to calibrate the appropriate level of substantiation for different users.

Third, any new framework should also consider how effective a predictive analytics program must be before clinical or patient use. As noted above, physician performance might be a sensible barometer. For example, an important milestone occurred in January 2017, when the FDA cleared the first product that relies on machine learning. The Arterys Cardio DL uses a self-teaching neural network to measure blood flow to the heart and thus aid cardiologists in diagnosing heart problems.¹⁰⁷ Arterys produced results on par with trained cardiologists, but took only 15 seconds to accomplish what might take a cardiologist well over 30 minutes.¹⁰⁸ Likewise, as discussed above, the deep neural network developed at Stanford to diagnose skin cancers produced results on par with dermatologists.¹⁰⁹ It makes little sense to prevent the use of such programs by practitioners for clinical decision support. The tougher question is whether these results justify use by patients and lay users. Moreover, even use by practitioners must account for the risk of automation bias and other realities of human-computer interaction.¹¹⁰ Finally, any new paradigm for predictive analytics must also contemplate what barometers might be appropriate for software functions that have no comparable physician counterparts and thus lack a baseline for validation.

A variety of different frameworks might accommodate these principles. For example, an FDA-led system for granting conditional approvals to predictive analytics programs and then requiring robust post-marketing studies and pre-set reevaluation timelines might generate reliable clinical evidence without inordinately delaying

special-review-procurement-procedures-related-utmdacc-oncology-expert-advis.pdf
[<https://perma.cc/KVH6-JNC7>].

¹⁰⁷ Letter from Robert Ochs, Dir., Div. of Radiological Health, U.S. Food & Drug Admin., to Ms. Golnaz Moeini, Dir. of Quality and Regulatory, Arterys Inc, 19, (Jan. 5, 2017), https://www.accessdata.fda.gov/cdrh_docs/pdf16/K163253.pdf [<https://perma.cc/6BPK-T6E6>].

¹⁰⁸ *Id.*; Bernard Marr, *First FDA Approval for Clinical Cloud-Based Deep Learning in Healthcare*, FORBES (Jan. 20, 2017), <https://www.forbes.com/sites/bernardmarr/2017/01/20/first-fda-approval-for-clinical-cloud-based-deep-learning-in-healthcare/#76a2889d161c> [<https://perma.cc/7MDQ-3TWM>].

¹⁰⁹ See Esteva et al., *supra* note 81, at 117.

¹¹⁰ Cortez, *Mobile Health Revolution*, *supra* note 3, at 1187-88.

market entry.¹¹¹ The agency might even rely on validation by third parties with more expertise in predictive analytics and other complex forms of software. Again, in June 2017, the agency announced a new “Digital Health Program” that would use non-governmental “certifiers” to screen digital health firms (rather than products), as well as shift the FDA’s expectations for data on safety and efficacy from the pre-market to post-market phase.¹¹² Forthcoming, then, are a slew of guidances from the agency explaining how these new ideas will be implemented.¹¹³

In the meantime, predictive analytics in health care will continue to develop as an important new tool for diagnosing and treating patients. But it will not reach its full potential without reliable ways to substantiate their claimed functions. Currently, without clear regulatory oversight, different types of non-government intermediaries are stepping in to evaluate these technologies, such as venture capital firms with experience in the health care sector, hospital systems, app review websites, and health insurers.¹¹⁴ Although each can screen and evaluate a limited number of these technologies, often relying on published literature for substantiation, they are at best short-term solutions for an industry that promises to expand dramatically in the coming years.

¹¹¹ Cortez et al., *supra* note 58, at 377.

¹¹² U.S. Food & Drug Admin., FDASIA HEALTH IT REP., *supra* note 59.

¹¹³ See FDA, Fostering Medical Innovation: A Plan for Digital Health Devices; Software Precertification Pilot Program, 82 Fed. Reg. 35,216 (July 28, 2017).

¹¹⁴ Cortez, *The Evolving Law*, *supra* note 89.

